

Movie Sales Prediction

Problem Description

Movies' meta-information and the sales are collected over the last five years.

The students are required to predict the two-week sales for the upcoming movies.

Evaluation Metric

Mean squared error

Data Source

The meta-information for both the past and the upcoming movies will be provided. The meta-information includes movies' title, type, developer, actor, director, popularity, and etc. Some of the attributes are not accessible for the upcoming movies.

The auxiliary data for the period up to October is allowed. The data will be available soon.

Supplement

The "totalBox" in "box.csv" stands for the total sales until present. "totalBox" may not equal to the summation of "dailyBox". Moreover, "boxtotal" in "metadata.csv" may not exactly equal to "totalBox" in "box.csv".

Our prediction objective is formally defined as "The summation of first 14-day dailyBox".

电影上映后每日相关信息数据 (box.csv) 中, totalBox 为截止到当前上映日期的票房总和, 该字段数值跟 dailyBox 字段的累加值并非严格一致, 仅供参考。同时, metadata.csv 中 boxtotal 数值跟 box.csv 中的 totalBox 也可能存在一定偏差。

最终预测结果评估，以电影日票房数据的前 14 天的和为准。

不同数据表中的电影名称可能存在一定的偏差，需要通过模糊匹配连接。